The Report Committee for Yiyi Chu
Certifies that this is the approved version of the following report:

Bayesian Mediation Analysis for Partially Clustered Designs

APPROVED BY
SUPERVISING COMMITTEE:

Supervisor:

Susan N. Beretvas

Michael Daniels

Bayesian Mediation Analysis for Partially Clustered Designs

by

Yiyi Chu, B.S.; M.A.

Report

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Master of Science in Statistics

The University of Texas at Austin

May 2013

Dedication

Dedicated to my loving parents and my fiancé *Fei Yan.*

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Tasha Beretvas. She has offered me steady guidance and unwavering support throughout my graduate study as well as my graduate career. She will always be a role model for me for her broad and in-depth knowledge, enthusiasm in research and teaching, devotion to her students, and grace. It has been an incredible journey learning from and working with her, and I am very grateful for her time, efforts and care.

Moreover, I would like to thank my committee member: Dr. Michael Daniels for his insightful advice, encouragement and the Bayesian class he taught me. His broad and in-depth knowledge in biostatistics and enthusiasm in research have inspired me and broadened my vision. I would also like to give my thanks to Dr. Peter Mueller for his valuable advice on my future PhD study, and the great MCMC course he taught me. I am always grateful for his kindness, patience and encouragement.

I would like to thank Dr. Deanna Staskel, whom I have worked with as her Teaching Assistant for two years, for her trust and support throughout my years at UT. My thanks also go to Dr. Mary Parker for teaching and helping me build a solid foundation in mathematical statistics.

I really appreciate everyone who graciously provided their expertise, time and resources for my graduate research in statistics, including Qingchao Meng, Baowei Xu and Qiong Zeng, who introduced me to this fascinating area of statistics and provided me with kindness help and suggestions; Youmin Wang, Zhan Shi, Ian Hembry and Li Wang, who generously allowed me to finish my simulation studies on their desktops.

My joyful time at The University of Texas at Austin was made largely due to the many friends that became a part of my life. They shared my excitement and anxiety and always cheered me up to move forward. I would like to acknowledge some of them: Wenyuan Li, Zhixing Zhang, Li Wang, Wanyi Wang, Ligang Long, and Zhan Shi.

Finally, I am indebted to my loving parents for their endless love, profound understanding and unwavering support of my pursuits. My deep gratitude goes to my fiancé, Fei Yan, for his love, encouragement and always being there for me. Thank you for coming into my life and giving me joy.

Abstract

Bayesian Mediation Analysis for Partially Clustered Designs

Yiyi Chu, M.S.Stat

The University of Texas at Austin, 2013

Supervisor: Susan. N. Beretvas

Partially clustered design is common in medicine, social sciences, intervention and psychological research. With some participants clustered and others not, the structure of partially clustering data is not parallel. Despite its common occurrence in practice, limited attention has been given regarding the evaluation of intervention effects in partially clustered data. Mediation analysis is used to identify the mechanism underlying the relationship between an independent variable and a dependent variable via a mediator variable. While most of the literature is focused on conventional frequentist mediation models, no research has studied a Bayesian mediation model in the context of a partially clustered design yet. Therefore, the primary objectives of this paper are to address conceptual considerations in estimating the mediation effects in the partially clustered randomized designs, and to examine the performances of the proposed model using both simulated data and real data from the Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). A small-scale simulation study was also conducted and the results indicate that under large sample sizes, negligible relative parameter bias was found in the Bayesian estimates of the indirect effects and of covariance between the

components of the indirect effect. Coverage rates for the 95% credible interval for these two estimates were found to be close to the nominal level. These results supported use of the proposed Bayesian model for partially clustered mediation in conditions when the sample size is moderately large.

Table of Contents

List of Tables

List of Figures

INTRODUCTION

Mediation hypotheses are commonly tested in social sciences, psychology and intervention research. A mediation model is used to identify the mechanism underlying the relationship between an independent variable and a dependent variable via a mediator variable, where the independent variable is hypothesized to influence the mediator, which subsequently influences the dependent variable. Substantial research has been conducted that has focused on single-level mediation analysis (see, for example, Pituch & Stapleton, 2008; Shrout & Bolger, 2002; MacKinnon, Lockwood, Hoffman, West, & Sheets, 2002). This basic model can be easily extended to provide multilevel mediation models that are appropriate for handling conditions in which the data are hierarchical or clustered (Krull & MacKinnon, 2001; Bauer, Preacher, & Gil, 2006; MacKinnon, Fairchild, & Fritz, 2006; Yuan & MacKinnon, 2009). While most of the educational and social science literature has focused on conventional frequentist approaches to estimating these mediation models, two recent articles have studied the mediation model from a Bayesian perspective (Yuan & MacKinnon, 2009; Daniels, Roy, Kim, Hogan, & Perri, 2012).

A Bayesian approach provides a useful and flexible alternative to the frequentist approach for estimating and formulating mediation models. By allowing an explicit use of external evidence from previous studies or some related resources (prior information) to be combined with the data collected in a current study, use of a Bayesian approach yields updated posterior distributions of the parameters of interest. In addition, the Bayesian framework provides a convenient framework for handling hierarchical data (Yuan & MacKinnon, 2009). Compared with conventional frequentist analyses such as maximum likelihood-based procedures, use of Monte Carlo Markov Chain (MCMC) estimation can simplify computation, and makes the estimation more efficient.

Furthermore, unlike when using conventional frequentist estimation procedures, use of Bayesian estimation does not rely on large-sample approximations (Robert, 2007).

In addition to exploring the use of Bayesian estimation for multilevel mediation models, the current study is also focused on extending mediation models for scenarios in which data are partially clustered. Partially clustered designs are encountered in medicine, social sciences, intervention and psychological research (Bauer, Sterba, & Hallfors, 2008). With some participants clustered in treatment groups and others not, the variance structures of partially clustered data cannot be assumed parallel. And despite its common occurrence in applied research, limited attention has been given regarding the evaluation of intervention effects in the partially clustered data. To date, seven articles have been found that have studied methods for estimating intervention effects in partially clustered intervention studies (see Hoover, 2002; Lee & Thompson, 2005; Myers, DiCecco, & Lorch, 1981; Roberts & Roberts, 2005; Bauer, Sterba, & Hallfors, 2008; Baldwin, Bauer, Stice & Rohde, 2011; Baldwin & Fellingham, 2012), and only one of these studies used Bayesian methods for estimating treatment effects. No research has been found that has explored estimation of a mediation model in the context of partially clustered data.

In this paper, we propose a Bayesian formulation of a multilevel mediation model to test mediation effects in the context of partially clustered design data. The primary objectives are 1) to address conceptual methodological considerations in estimating the mediation effects in partially clustered randomized designs; and 2) to examine the performances of the proposed model using both simulated and real data. The outline of this paper is as follows. First, we include a brief review of Bayesian estimation and Gibbs sampling. Second, we review mediation models' estimation from both conventional and Bayesian perspectives as well as common conventional frequentist statistical methods that are used when assessing the indirect effects in these models. Third, we introduce the partially clustered design, and discuss a general framework for modeling this type of data. Fourth, we propose a specific model to conduct multilevel mediation analysis for data sampled from a partially clustered randomized design, and discuss how to estimate this model using Bayesian estimation. Despite our focus on a basic hierarchical model, here, for simplicity's sake, it is relatively straightforward to extend the approaches and considerations addressed here to more complicated models within similar contexts. Fifth, we test estimation of the model's key parameters of interest by using both simulated data and real data from The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K). Finally, we discuss the results and detail some potential future extensions.

BAYESIAN INFERENCE

Frequentist statistical inference treats a parameter as an unknown, single fixed value. In contrast, Bayesian inference uses probabilities (or a probability distribution) to measure uncertainty about the unknown parameter, and thus considers parameters as random entities. The mechanism at the foundation of Bayesian inference is summarized by Bayes theorem:

$$\mathrm{p}(\theta|\mathrm{data}) = \frac{p(\theta)p(\mathrm{data}|\theta)}{p(\mathrm{data})}, \tag{2-1}$$

where θ represents the unknown parameter, $p(\mathrm{data}|\theta)$ is the probability distribution of data given θ (the likelihood function); $p(\theta)$ is the probability distribution of θ quantifying the knowledge of θ prior to seeing the data (the prior distribution); $p(\mathrm{data}) = \int(\theta)p(\mathrm{data}|\theta)d\theta$, which is the marginal density of the data after integrating out θ. Because $p(\mathrm{data})$ is a normalizing constant, Bayes theorem is often written as: $\mathrm{p}(\theta|\mathrm{data}) \propto p(\theta)p(\mathrm{data}|\theta)$. The resulting probability density $\mathrm{p}(\theta|\mathrm{data})$ is known as the posterior distribution because it combines the prior information before observing the data with the information from the data that have been observed. From the posterior density, several commonly used summary statistics can be calculated to draw inferences about θ. One point estimate of θ is its posterior mean, given by

$$\hat{\theta} = E(\theta|data) = \int \theta p(\theta|data)d\theta. \tag{2-2}$$

If the posterior distribution is skewed, the posterior distribution's mode or median could provide an alternative point estimate of θ. Additional important point estimates describing the parameter's posterior distribution include the posterior variance:

$$\sigma_\theta^2 = \mathrm{Var}(\theta|\mathrm{data}) = \int \left(\theta - \hat{\theta}\right)^2 p(\theta|data)d\theta, \tag{2-3}$$

And the posterior standard deviation

$$\sigma_\theta = \sqrt{Var(\theta|data)} = \sqrt{\int \left(\theta - \hat{\theta}\right)^2 p(\theta|data)d\theta}. \quad (2\text{-}4)$$

which provide measures of the uncertainty in the parameter's estimation from a Bayesian perspective.

In addition to point estimates, Bayesian "credible intervals" can provide interval summaries that are somewhat analogous to frequentist confidence intervals although interpretation of a credible interval's limits is more straightforward. The 95% credible interval is quantified as $(q_{0.025}, q_{0.975})$, where $q_{0.025}$ and $q_{0.975}$ represents the 2.5% and 97.5% quantile of the posterior distribution, respectively.

If the posterior distribution $p(\theta|data)$ follows a specific distributional form, it is easy to calculate the summary statistics including posterior mean, variance and credible interval. However, this rarely happens especially in the high-dimensional problems. The most common approach in such situations is to use MCMC methods to get samples from the posterior distribution, and then use sample mean, variance, and quantile to estimate the posterior summary statistics based on the posterior draws. Under MCMC framework, a sequence of correlated samples is generated, each of which is correlated with adjacent samples. As a result, thinning can be used if independent samples are desired (Gelman, Carlin, Stern, & Rubin, 2003). In addition, since the MCMC algorithm can take a large number of steps to reach the desired distribution, samples from the beginning of the Markov chain are often not used.

Gibbs sampling is a common MCMC algorithm for regression models. It uses the full conditional posterior distribution to update each parameter, assuming the others are known. Through iteratively repeating draws from the full conditional distributions, we end up with obtaining a sequence of values that approximate the joint distribution of the parameters of interest. By examining the sample mean, sample variance and 95% quantile interval on the basis of posterior draws, inferences can be made about the parameters of interest. When the full conditional density does not follow a known probability distribution, alternative sampling algorithms can be considered, such as Metropolis Hastings algorithm, slice sampling and rejection sampling (Gelman, Carlin, Stern & Rubin, 2003).

MEDIATION MODELS

MEDIATION MODELS FOR SINGLE LEVEL DATA

A single-level mediation model is suitable for unclustered data and can be expressed as follows:

$$\mathrm{M}_i = \beta_1 + \alpha X_i + e_{1i}, \tag{3-1}$$

$$\mathrm{Y}_i = \beta_2 + \beta M_i + \tau' X_i + e_{2i}, \tag{3-2}$$

where *Y*, *X*, and *M* are the values of dependent variable, independent variable, and mediating variable (mediator), respectively. α measures the relationship between the independent variable and mediating variable (path *α* in Figure 3.1). β represents the effects of mediating variable on the dependent variable after controlling for the effects of the independent variable (path *β* in Figure 3.1), and τ' represents the effects of independent variable X on the dependent variable Y after controlling for the mediator (path τ' in Figure 3.1). e_1 and e_2 are assumed to be normally distributed with mean of 0 and variances of σ_1^2 and σ_2^2, respectively. The mediation effect (indirect effect) is thus estimated by $\hat{\alpha}\hat{\beta}$ (MacKinnon & Dwyer, 1993).

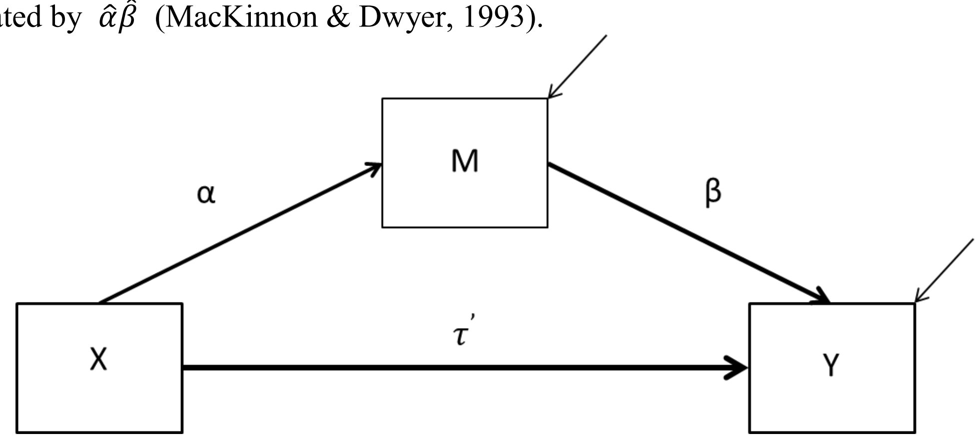

Figure 3.1 The Single-Level Mediation Model.

Conventional Estimation

Several methods are widely used to assess whether this mediation effect $\hat{\alpha}\hat{\beta}$ is significantly different from zero. One of the most often used tests is the *z* test. In this test, the mediation effect $\hat{\alpha}\hat{\beta}$ is divided by an estimate of its standard error (Sobel, 1982), which is calculated as

$$\hat{\sigma}_{\hat{\alpha}\hat{\beta}} = \sqrt{\hat{\sigma}_{\beta}^{2}\hat{\alpha}^{2} + \hat{\sigma}_{\alpha}^{2}\hat{\beta}^{2}} \tag{3-3}$$

where $\hat{\sigma}_{\alpha}^{2}$ and $\hat{\sigma}_{\beta}^{2}$ are the sampling variances of $\hat{\alpha}$ and $\hat{\beta}$ respectively. The ratio $\frac{\hat{\alpha}\hat{\beta}}{\hat{\sigma}_{\hat{\alpha}\hat{\beta}}}$ is then compared to critical values from the standard normal distribution, as z test assumes that the sampling distribution of the mediation effect is normal. However, the distribution of $\hat{\alpha}\hat{\beta}$ can be skewed and kurtotic (Yuan & MacKinnon, 2009). Therefore, the *z* test results in the use of normal approximation that is often not appropriate.

The empirical-*M* test has been proposed as an improved confidence interval method (MacKinnon, Lockwood, & Williams, 2004). Instead of assuming a normal sampling distribution for the product $\hat{\alpha}\hat{\beta}$, this test assumes that each of the parameters, $\hat{\alpha}$ and $\hat{\beta}$, has a normal distribution. To conduct the empirical-*M* test, the improved confidence interval is calculated by finding the upper and lower limit as:

$$\text{Upper limit} = \hat{\alpha}\hat{\beta} + C.V.up * \hat{\sigma}_{\hat{\alpha}\hat{\beta}}, \tag{3-4}$$

$$\text{Lower limit} = \hat{\alpha}\hat{\beta} + C.V.low * \hat{\sigma}_{\hat{\alpha}\hat{\beta}}, \tag{3-5}$$

where $\hat{\sigma}_{\hat{\alpha}\hat{\beta}}$ is the Sobel standard error as used in the *z* test in Equation (3-3). The asymmetric values for the critical values (*C.V. up* and *C.V. low*) are generated from computationally intensive simulations. The resulting confidence interval is not necessarily symmetric.

Another improved confidence interval method is obtained using bootstrap resampling, which does not require assumptions about the distribution of the statistic. Because the product $\hat{\alpha}\hat{\beta}$ does not follow a normal distribution, bootstrapping has been suggested as a recommended way of testing mediation effects, especially in studies with small samples (Bollen & Stine, 1990; Preacher & Hayes, 2004; Shrout & Bolger, 2002).

Bayesian Estimation

Bayesian inference starts with specifying priors for all the unknown parameters. In the context of single-level mediation analysis, unknown parameters θ= $\{\beta_1, \alpha, \beta_2, \beta, \tau', \sigma_1^2, \sigma_2^2\}$. Since there are two linear regression equations involved, normal priors are often used for the regression coefficients $\{\beta_1, \alpha, \beta_2, \beta, \tau'\}$. For the residual variances$\{\sigma_1^2, \sigma_2^2\}$, the commonly used prior distribution is inverse Gamma ($\sigma_1^2 \sim IG$ (a, b)). With regards to specifying the hyperparameter values for prior distribution, it depends on how much and what prior information researchers have (non-informative priors versus informative priors). If researchers want data to play a dominant role on the inference, a flat prior could be used. Specifically, large values like 10,000 can be assigned to the variance parameters of the normal priors to indicate strong uncertainty about the external information other than data. Small values like 0.001 can be assigned to a and b, so that the prior has a large variance and is very flat over the range. If historical data from previous studies are available, an informative prior can be employed to provide more resources outside the observed data (Gelman, Carlin, Stern & Rubin, 2003). For a detailed introduction to Bayesian prior specification, see Gelman *et al*., (2003).

For Equation (3-1) from single-level mediation model, posterior draws of $\beta_1, \alpha, \sigma_1^2$ can be obtained by implementing a Gibbs sampling procedure as follows, for the s^{th} draw:

1. Draw $\beta_1^{(s)}, \alpha^{(s)}$ from their full conditional distribution $p\ (\beta_1, \alpha | \sigma_1^{2(s-1)}, data)$
2. Draw $\sigma_1^{2(s)}$ from its full conditional distribution $\mathrm{p}\ (\sigma_1^2 | \beta_1^{(s)}, \alpha^{(s)}, data)$

Similarly, posterior draws of β_2, β, τ' and σ_2^2 can be obtained by fitting Equation (3-2) with specific priors and then implementing a Gibbs sampler.

Inferences for any functions of these parameters can be made once S posterior draws of $\{\beta_1, \alpha, \beta_2, \beta, \tau', \sigma_1^2, \sigma_2^2\}$ are obtained. Because our key parameter of interest in the mediation model is the indirect effect $\hat{\alpha}\hat{\beta}$, the posterior mean and variance can be calculated using the following:

$$\widehat{\alpha\beta} = \frac{1}{S}\sum_{s=1}^{S} \alpha^{(s)}\, \beta^{(s)}, \tag{3-6}$$

$$\mathrm{Var}(\alpha\beta|\mathrm{data}) = \frac{1}{S-1}\sum_{s=1}^{S} (\alpha^{(s)}\beta^{(s)} - \widehat{\alpha\beta})^2. \tag{3-7}$$

To use the Bayesian framework for estimation purposes only with the intention of conducting frequentist statistical hypothesis testing, the mediation effect can be tested against a value of zero using a 95% credible interval. If the value of zero is contained within the interval, then the inference would be that the parameter's value is not likely to differ from zero.

MEDIATION MODELS FOR TWO-LEVEL DATA

When the data are hierarchically structured, for example, in educational datasets in which students are clustered within schools and in longitudinal datasets where observations over time are clustered within subjects, cases cannot be assumed to independent. One option for handling dependent data involves use of a multilevel model. Multilevel mediation models are more complicated than single-level mediation both computationally and conceptually. Unlike the general single-level mediation model, there are several possible multilevel mediation models that are distinguished by the level at which each variable is measured. Krull and MacKinnon (2001) offered the L_X - L_M - L_Y notation where L_i represents the level at which variable *I* is measured. The most common configurations include the following scenarios in which: the independent (*X*) and mediator variables (*M*) are measured at level-2, while the outcome variable (*Y*) is measured at level 1 (i.e., 2-2-1); *X* is measured at level-2, while M and Y are measured at level 1 (2-1-1); all three variables are measured at level-1 (1-1-1). Our focus in this paper is on the 1-1-1 mediation model, which is written as follows, at level one, the outcome is modeled as a function of *X*:

$$M_{ij} = \beta_{2j} + \alpha_j X_{ij} + e_{2ij}, \tag{3-8}$$

and the distal outcome, *Y* is modeled as a function of both *X* and *M*:

$$Y_{ij} = \beta_{3j} + \beta_j M_{ij} + \tau_j' X_{ij} + e_{3ij}, \tag{3-9}$$

where *i* indexes the first level, and *j* represents the second level.

At level-2, the model for the coefficients in Equation (3-8) is as follows:

$$\beta_{2j} = \beta_2 + u_{1j}, \tag{3-10}$$

$$\alpha_j = \alpha + u_{2j}, \tag{3-11}$$

and the level-2 model for the coefficients in Equation (3-9) is:

$$\beta_{3j} = \beta_3 + u_{3j}, \tag{3-12}$$

$$\beta_j = \beta + u_{4j}, \tag{3-13}$$

$$\tau_j' = \tau' + u_{5j}. \tag{3-14}$$

where e_{2ij} and e_{3ij} are the level-1 residuals for *M* and *Y*, respectively. They are assumed to be independent and normally distributed such that:

$$e_{2ij} \sim N(0, \sigma_2^2) \tag{3-15}$$

$$e_{3ij} \sim N(0, \sigma_3^2) \tag{3-16}$$

β_{2j} and β_{3j} are random intercepts for Equation (3-8) and (3-9), respectively, and α_j, β_j and τ_j' are the random slopes. Specifically, α_j measures the effect of the independent variable on the mediator, and β_j represents the mediator's effect on the dependent variable, after adjusting for the independent variable; and τ_j' measures the direct effects of the independent variable on the outcome (after controlling for the mediator). The parameters α, β represent the average effect of the independent variable on the mediator, and the average effect of the mediator on *Y* after controlling for the independent variable, respectively. The level-2 residuals $\mu_j = (\mu_{1j}, \mu_{2j}, \mu_{3j}, \mu_{4j}, \mu_{5j})^T$ are typically assumed to follow a multivariate normal distribution with mean vector of zeroes and a 5 x 5 variance-covariance matrix:

$$\mu_j \sim N(\mathbf{0}, \Sigma). \tag{3-17}$$

The average indirect effect of independent variable X_{ij} on dependent variable Y_{ij} through the mediator variable M_{ij} in the 1-1-1 multilevel mediation model is thus given by

$$\text{ab} = \text{E}\left(\alpha_j \beta_j\right) = \alpha\beta + \sigma_{\alpha_j \beta_j}, \tag{3-18}$$

where $\sigma_{\alpha_J \beta_J}$ represents the covariance between α_j and β_j.

Conventional Estimation

When the dependence inherent in multilevel data is ignored, by using, for example, ordinary least squares estimation, then biased estimates will result (Yuan & MacKinnon, 2009). Instead, maximum likelihood estimation and empirical Bayes methods are usually used with clustered data. In the context of multilevel mediation that involves two regression equations at the first level (see Equation (3-8) and (3-9)), it becomes more challenging to estimate the covariance components between pairs of random effects. Bauer et al. (2006) proposed a method that uses a selection variable to estimate two level-1 regression equations simultaneously as follows:

$$Z_{ij} = I_{Y_{ij}}\left(\beta_{3j} + \beta_j M_{ij} + \tau_j' X_{ij} + e_{3ij}\right) + (1 - I_{Y_{ij}})\left(\beta_{2j} + \alpha_j X_{ij} + e_{2ij}\right), \quad \text{(3-19)}$$

where $I_{Y_{ij}}$ denotes a selection variable such that when $I_{Y_{ij}}$=1, then the outcome Z_{ij}, represents Y_{ij}, otherwise Z_{ij} represents M_{ij}. Fitting this model in Equation (3-19) permits simultaneous estimation of both models and provides consistent estimates of variance and covariance components (Bauer, Preacher & Gil, 2006). One weakness of conventional frequentist, maximum likelihood-based estimation, however, is that the procedure can yield extreme values for covariance parameter estimates when sample sizes are small. This is because small sample sizes fail to provide enough information about covariance parameters in the data (Baldwin & Fellingham, 2011).

Bayesian Estimation

Use of Bayesian methods provides a more straightforward way to handle some of the complexities encountered when estimating a 1-1-1 multilevel mediation model. First, multilevel or hierarchical models fit naturally into the Bayesian framework, as parameters are considered to be random rather than fixed. Without imposing the restriction of estimating the full model simultaneously and any additional adjustments, Bayesian modeling automatically takes into account the uncertainty associated with covariance parameters. Moreover, as discussed earlier, no asymptotic approximations are necessarily needed when using Bayesian methods, which are computationally more efficient as compared with conventional non-parametric methods, such as bootstrapping (Yuan & MacKinnon, 2009).

In terms of the prior specification for multilevel mediation models, non-informative normal priors with extremely larger variance values or informative priors elicited from historical studies can be assigned to the coefficient parameters. For the unknown residual variances parameters, either non-informative flat priors, such as uniform priors and vague inverse gamma priors, or informative priors on the basis of historical information can be used. By implementing a Gibbs sampler as discussed in the single-level mediation model, we could obtain posterior draws of α, β, and $\sigma_{\alpha\beta}$ from the dependent Markov chain, and thus obtain estimates of the average mediation effect by calculating the following

$$ab^{(s)} = \alpha^{(s)}\beta^{(s)} + \sigma_{\alpha_j\beta_j}^{(s)} \; s = 1, 2, \dots\dots, S,$$

where $\alpha^{(s)}$, $\beta^{(s)}$and $\sigma_{\alpha_j\beta_j}^{(s)}$ represent the s^{th} posterior samples from the realization distribution. The posterior mean and variance of the average mediation effects are, respectively, as follows:

$$\widehat{ab} = E(ab|data) = \frac{1}{S}\sum_{s=1}^{S} ab^{(s)}, \tag{3-20}$$

$$\mathrm{Var(ab|data)} = \frac{1}{S-1}\sum_{s=1}^{S}(ab^{(s)} - \widehat{ab})^2. \tag{3-21}$$

To examine whether the average mediation effect is significantly different from zero, a 95% credible interval can be obtained in the same manner as discussed for the single-level mediation model.

PARTIALLY CLUSTERED MODELS

Traditional hierarchical models for fully clustered designs, have parallel structures such that clustering occurs in the same way for all study conditions. In contrast, a nonparallel structure is observed in partially clustered designs, such that some study conditions involve clustering of, say, individuals within groups, while other conditions do not. For example, a study that compares married couples with single individuals can be considered as a partially nested study, as there are actually two sets of participants involved in the study—clustered participants within dyads (married couples) and unclustered participants (single individuals) (Baldwin, Bauer, Stice & Rohde, 2011). Alternatively, it might be the case that one set of individuals are randomly assigned to treatment groups while randomly assigned control participants are not assigned to groups. The treatment group participants are then members of groups while the control participants are not. In this kind of partially clustered design dataset, modeling of a cluster-level variance is only needed for participants' data in some but not all study conditions. A simple model for handling partially clustered data was suggested by (Bauer, Sterba & Hallfors, 2008).

At level-1, the model was as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j}X_{ij} + e_{1ij}, \tag{4-1}$$

And at level-2, the model was:

$$\beta_{0j} = \gamma_{00}, \tag{4-2}$$

$$\beta_{1j} = \gamma_{10} + u_{1j}, \tag{4-3}$$

where the level-1 and level-2 residuals e_{1ij} and u_{1j} are assumed to be independent and normally distributed such that

$$e_{1ij} \sim \mathrm{N}\,(0, \sigma^2),$$

and

$$u_{1j} \sim N\,(0, \tau_{11}).$$

In this model, X_{ij} denotes a dichotomous variable, where X_{ij}=1 represents the treatment condition (experienced in clusters), and X_{ij}=0 the control condition (not experienced in clusters). β_{0j} denotes the group mean for control arm subjects (for whom $X_{ij} = 0$). Because each control participant is the sole member for every control "group", each group "mean" is assumed the same. No group- or cluster-level residual is modeled for unclustered control condition scenarios. To summarize, the model for unclustered control condition given $X_{ij} = 0$. is given by:

$$Y_{ij}\big|\big(X_{ij} = 0\big) = \gamma_{00} + e_{1ij}, \tag{4-4}$$

where the variation in the outcome variable, Y_{ij}, for control participant *i* is decomposed into an overall control condition mean $\gamma_{00}$and an individual-specific residual e_{1ij}. β_{1j} represents the difference between the mean for clustered treatment group *j* and the overall mean for control participants. Because the treatment condition involves clustering within treatment groups, the group mean is allowed to vary across treatment groups by assigning a random component u_{1j} to β_{1j} in Equation (4-3).The model for clustered treatment condition given $X_{ij} = 1$ is then given by:

$$Y_{ij}\big|\big(X_{ij} = 1\big) = \gamma_{00} + \gamma_{10} + u_{1j} + e_{1ij}. \tag{4-5}$$

where e_{1ij} here measures the individual participants' treatment effect differences within a treatment group, and u_{1j} captures differences in the average treatment effect across groups.

Due to the nonparallel variance structure in partially clustered data, the variance of the outcome variable differs by study condition. For participants in the (unclustered) control arm, the variance is given by:

$$\text{Var}\left(Y_i \middle| X_{ij} = 0\right) = \sigma^2. \tag{4-6}$$

while the variance for participants in the (clustered) treatment arm is

$$\text{Var}\left(Y_{ij} \middle| X_{ij} = 1\right) = \sigma^2 + \tau_{11}. \tag{4-7}$$

Roberts and Roberts (2005) noted that it is more reasonable to allow the level-1 residual variance σ^2 to differ across the treatment versus control arm, because the clustering structure tends to increase the within-group variability as compared with that found in the control arm. Thus, the model assumption regarding level-1 residuals should be modified to be heteroscedastic such that the level one residuals' variance for control participants, $\sigma^2_{control}$, is modeled as different from that for the treated participants' level one residuals' variance, $\sigma^2_{treatment}$.

ESTIMATION CHALLENGES FROM CONVENTIONAL APPROACH

Partially clustered data can involve small sample sizes and nonparallel data structures, which makes the partially clustered data's model more complex to estimate. As summarized in Bauer et al. (2008), two major estimation challenges have been encountered when estimating such kind of model by conventional likelihood methods. First, underestimation of the standard errors tends to occur. Under the framework of likelihood methods, fixed effects estimation highly depends on the values of the variance and covariance parameter estimates. Because unbalanced data with small sample sizes tends to provide insufficient information to estimate the variance parameters well, this often results in biased estimates of variance and/or covariance components as well as deflated standard errors and too narrow confidence intervals for the fixed effects. Second, since the distributional forms for the parameters of interest are unknown, p-values and confidence intervals are difficult to obtain. To deal with the first problem, Kenward and Roger (1997, 2009) proposed to use an adjustment to standard errors of the fixed effects, so that uncertainties associated with estimating the variance and covariance components could be taken into account through inflation of the standard errors. For the second inferential challenge, the *t*-distribution is often assumed to approximate the sampling distribution of fixed effects components, with degrees of freedom approximated by the Satterthwaite adjustment (Kenward & Roger, 1997; Baldwin & Fellingham, 2012). The Kenward and Roger methods proved to provide appropriate adjustments with the exception of variance components estimated at the boundary. However, when the sample size and especially when the number of clusters is small, limited information can be acquired from data to estimate the variance components, leading to boundary problems for cluster variance estimates. In such situations, adjustments to standard errors and degrees of freedom do not work. Therefore, the uncertainty in estimation of the variance parameters is still a potential issue in fitting and estimating the partially clustered models, where small number of clusters and small sample sizes are encountered, and variance structures are complex to deal with (Baldwin & Fellingham, 2012).

BAYESIAN ESTIMATION

The inferential challenges that likelihood estimation struggles to handle can be accommodated by using Bayesian estimation. As a probability distribution, the prior distribution specifies the uncertainties in unknown parameters before considering the data. The posterior distribution, which combines the prior and data information, then automatically allows for these uncertainties in the parameters, without requiring additional adjustments. Bayesian methods also help deal with the boundary problems that are met in maximum likelihood estimation through the use of prior information (Baldwin & Fellingham, 2012). Priors play dominant roles when the sample sizes are small. In such cases, the posterior estimates will be "shrunken" to the prior mean. This typical shrinkage behavior underscores the importance of choosing appropriate priors, as wrong inferences would be made if posterior estimates are shrunk to implausible values. In addition, no approximation needs to be imposed when using the Bayesian framework. Unlike traditional likelihood methods that assume a *t*-distribution for fixed effects estimates and requires the Satterthwaite adjustment to adjust for unknown degrees of freedom, the Bayesian posterior distributions are the true probability distributions for the parameters of interest, and researchers can make inferences directly based on them. Finally, as discussed in the section of mediation models, multilevel models are more naturally handled under the Bayesian paradigm and Bayesian methods can be especially advantageous for small-sample data. Lots of well-developed MCMC algorithms are widely used for fitting the more complicated hierarchical models.

MULTILEVEL MEDIATION MODELS IN PARTIALLY CLUSTERED DESIGN

Despite the comparatively little attention that has been paid to the evaluation of treatment effects in partially clustered designs versus fully clustered designs, the partially clustered design is actually quite common among randomized experiments in practice (Bauer, Sterba & Hallfors, 2008). As noted by Bauer et al. (2008), 32% of the sampled studies (30 out of 94) from four representative clinical research and public health journals were found to use partially clustered designs, and were found to be more prevalent than fully clustered designs (found in 26 out of the 94 sampled studies). As another commonly studied and widely used statistical model in clinical and intervention research, the mediation model helps understand the mechanisms through which the independent variable affects the outcome variable. How to fit the multilevel mediation model and then evaluate the mediation effects in group-randomized (fully-clustered) design is well documented. To our knowledge, however, no research has been conducted on the multilevel mediation analysis in partially clustered designs. Given the characteristics of this type of design and of the complexities of testing mediation effects, we propose the partially clustered multilevel mediation model as shown below. For notational simplicity, we use a simple two-level partially clustered mediation model. Extensions to models that include additional, higher levels or more predictor variables are straightforward.

Let i denote the index units of the first level and j denote the index units of the second level. A multilevel 1-1-1 mediation model with one treatment arm involving individuals clustered within groups and a control arm in which participants are not clustered in groups is given by

at level-1:

$$M_{ij} = \beta_1 + \alpha_j X_{ij} + e_{1ij}, \tag{5-1}$$

$$Y_{ij} = \beta_2 + \beta_{3j} X_{ij} + \beta M_{ij} + u_{4j} M_{ij} X_{ij} + e_{2ij}, \tag{5-2}$$

At level-2:

$$\beta_1 = \gamma_{10}, \tag{5-3}$$

$$\alpha_j = \alpha + u_{1j}, \tag{5-4}$$

$$\beta_2 = \gamma_{20}, \tag{5-5}$$

$$\beta_{3j} = \gamma_{30} + u_{2j}, \tag{5-6}$$

X_{ij} denotes a dichotomous variable, where X_{ij}=1 represents the clustered treatment condition, and X_i=0 represents the unclustered control condition. Residuals e_{1ij} and e_{2ij} are assumed to be independent and normally distributed as follows:

$$e_{1ij} \sim N\ (0, \sigma_1^2),$$
$$e_{2ij} \sim N\ (0, \sigma_2^2).$$

The parameters β_1 and β_2 represent fixed intercepts that are fixed across clusters, and $\alpha_j, \beta_{3j}, \beta_j$ are random slopes. Specifically, β_1 indicates the mean value of the mediator variable for subjects in the control condition. Since there is no clustering structure for the control arm, no random component is needed for β_1. Similarly, the intercept for Y, β_2, is treated as fixed. α_j represents the treatment effect by measuring the difference between the mean for treatment group j and the overall mean for the control arm; β_{3j} measures the direct treatment effects on the outcome variable Y, after controlling for the mediator. Given the hierarchical structure in the treatment arm, random components are included in the level-2 equations for α_j and β_{3j}, respectively. Whether a level-2 random effect for M_{ij} should be included depends on whether subject i is in the treatment or control arm of the study, which is indicated as $u_{4j}M_{ij}X_{ij}$ here. Thus, the independent variable X_{ij} is used, which equals 1 when subject i is in the clustered arm, and equals 0 when in the unclustered control arm. Therefore, $\beta + \mu_{4j}$, which is, β_j, represents the mediating effect after adjusting for X_{ij} in the clustered treatment arm, and is modeled to vary across treatment groups. In the second-level model, parameters α and γ_{30} represent average slopes in the clustered treatment condition. Second-level residuals $\mu_j = (\mu_{1j}, \mu_{3j}, \mu_{4j})^T$ follow a multivariate normal distribution with mean vector of zeroes and a 3 x 3 variance-covariance matrix, which can be written as

$$u_j \sim N\,(0, \Sigma).$$

We permit heteroscedasticity in this model by allowing the level-1 residual variances to differ across the study arms. This proposed model seems to be consistent with the partially clustered design, and to reflect all those characteristics that mediation models have. In the next two sections, a real-data example and a simulation study are presented to illustrate and assess Bayesian estimation of the partially clustered mediation model.

RESULTS FROM THE ECLS-K STUDY

ECLS-K Study Description

An important data source: The Early Childhood Longitudinal Study, Kindergarten Class of 1998-99 (ECLS-K) was used to demonstrate and compare use of conventional likelihood-based estimation with Bayesian estimation of the multilevel mediation model for partially clustered designs. Independent variable X represents exposure to one of the two conditions: the full-day kindergarten program (n=408), and the half-day kindergarten program (n=408). Our analytic sample includes 68 schools that offer full-day kindergarten programs, with 6 randomly selected students nested within each school to make the clustered condition sample (cluster size =6, number of clusters=68). 408 students from 408 different schools (i.e., one student per school) that offer half-day kindergarten programs were selected into the unclustered condition. The mediator variable M was a measure of the math performance in the fall of kindergarten (1998), and the dependent variable Y was a measure on mathematics scores in the spring of kindergarten (1999). We hypothesized that full-day kindergarten program could benefit students' math performance in fall kindergarten, which in turn leads to improvement of math performance in spring kindergarten. For simplicity, we assume that there are no other covariates in the level-1 or level-2.

Although not designed as partially clustered data, this ECLS-K analytic sample was selected to have an exact partially clustered structure, with the intervention condition (full-day kindergarten) involving clustered data, and the control condition (half-day kindergarten) entailing unclustered student data. The partially clustered multilevel mediation model discussed above was applied to this ECLS-K data. Both likelihood-based estimation procedures and Bayesian methods were used to analyze this real data set.

LIKELIHOOD-BASED ESTIMATION RESULTS

We conducted a conventional partially clustered multilevel mediation analysis by estimating the two level-1 equations simultaneously using (3-19) in R package "nlme"(Bauer, Preacher & Gil, 2006). Unfortunately, the solution did not converge. As noted in Bauer et al. (2006), nonconvergence or boundary solutions are not uncommon problems when fitting a complicated model using likelihood-based estimation. The authors examined the rate of nonconvergence, boundary solutions and unconstrained solutions in their simulation study, and found that when there were 50 level-2 units with four subjects in each unit, more than half of the fitted models failed to reach convergence (54.7%). When the number of clusters was increased to 100, however, the nonconvergence rates decreased dramatically (down to 16.52%). These estimation problems imply that level two sample size is of great importance for estimating these kinds of models. For the current ECLS-K dataset, we have a relatively small cluster size (N1 = 6), and a moderate number of level-2 units (N2 = 68), suggesting nonconvergence is a reasonable result.

BAYESIAN RESULTS

We assumed that the level-2 random slope that represents the treatment effect on the mediator variable M_{ij}, α_j, and the parameter, β_j, that represents the adjusted mediating effect on the outcome variable were correlated, and followed a bivariate normal distribution with a mean vector $\binom{\alpha}{\beta}$, and a variance-covariance matrix $\sum$ that can be written as

$$\sum = \begin{pmatrix} \sigma_{\alpha_j}^2 & \sigma_{\alpha_j\beta_j} \\ \sigma_{\alpha_j\beta_j} & \sigma_{\beta_j}^2 \end{pmatrix}.$$

The level-2 random slope that represents the direct effect of X_{ij}, β_{3j}, was assumed to be independent of α_j and β_j. This assumption corresponds to what was assumed in previous related research (see Yuan & MacKinnon, 2009).

Given that the total sample size was moderately large (n=408) and that no historical information was available to use as prior information, noninformative priors were used. Specifically, vague normal priors were assigned to the regression parameters:

$$\mathrm{p}(\beta_1) \sim N\,(0, 10^6), \tag{6-1}$$

$$\mathrm{p}(\beta_2) \sim N\,(0, 10^6), \tag{6-2}$$

$$\mathrm{p}(\gamma_{30}) \sim N\,(0, 10^6), \tag{6-3}$$

$$\mathrm{p}(\alpha_j, \beta_j) \sim Bivariate\ Normal\ (\binom{\alpha}{\beta}, \sum), \tag{6-4}$$

where a scaled inverse Wishart prior was used for the variance-covariance matrix, $\sum$, because of the conditionally conjugate features. The inverse Wishart distribution is characterized by two hyperparmaters, the degree of freedom, denoted as ρ, and the scale matrix, denoted as A. A commonly employed diffuse prior for ρ is to make ρ equal to the dimension of the random coefficient vector, which is 2 here. In terms of the specification of hyperparameter A, matrix structures with identity matrix and scalar multiples could be considered (Gelman & Hill, 2007). Here an identity matrix of $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ was set as the scale matrix on the basis of the discussions on scaled inverse Wishart priors in Gelman & Hill, (2007). And

$$p(\alpha) \sim N\,(0, 10^6), \tag{6-5}$$

$$p(\beta) \sim N\,(0, 10^6). \tag{6-6}$$

Vague gamma priors were assigned for the level-1 variance parameters:

$$p(\sigma_1^{-2}) \sim Gamma\,(0.001, 0.001), \tag{6-7}$$

$$p(\sigma_2^{-2}) \sim Gamma\,(0.001, 0.001). \tag{6-8}$$

For the variance of β_{3j}, a vague uniform prior was used:

$$p(\sigma_{\beta 3j}) \sim Unif(0,100), \tag{6-9}$$

where $\sigma_{\beta 3j}$ represents the standard deviation of β_{3j}. We used 50,000 posterior draws to make Bayesian inferences with the first 10,000 iterations burnt in. Results are summarized in Table 6.1 and Figure 6.1.

As shown in the Table 6.1, the 95% credible interval for the indirect effect (*ab*) was (-1.280, 2.345), included zero. This implies the indirect effect is not significantly different from zero. There was a not significant covariance detected between α_j and β_j, given the 95% credible interval for $\sigma_{\alpha_j\beta_j}$ did include zero (-0.764, 0.071).

A graphical convergence diagnosis of the MCMC algorithms was presented in the Figure 6.1, which shows the trace plots of the posterior samples for the indirect effect (*ab*) and the covariance between α_j and β_j ($\sigma_{\alpha_j\beta_j}$). The convergence of the indirect effect (*ab*) looked good; Towards the end of the chain, however, the trace associated with the covariance between α_j and β_j ($\sigma_{\alpha_j\beta_j}$) became not that smooth, indicating a potential mixing issue there. A formal convergence diagnostic test, the Geweke's convergence diagnostic based on a test for equality of the means of the first and last part of a Markov chain was also conducted (Geweke, 1992). Convergence is supported if the two means do not differ significantly. Geweke's statistic is assumed to follow an asymptotically standard normal distribution. The Geweke's statistic values for the indirect effect (ab) and the covariance between α_j and β_j equaled 1.574 and -1.207, respectively. Thus, neither statistic differed significantly from zero supporting the conclusion of satisfactory convergence.

Because a real data analysis does not tell what true values of the parameters should be, the accuracy of the corresponding parameter estimates based on the proposed statistical model cannot be evaluated. In this case, a simulation study is needed such that an assessment of the model performance can be conducted using "true" parameter values against which to test estimates. In the next chapter, we present a small-scale simulation study designed to evaluate estimation of the partially clustered mediation model.

Table 6.1 Bayesian Estimates of Parameters in the Partially Clustered Mediation Model

	Mean	**Median**	**SD**	**2.50%**	**97.50%**
ab	0.566	0.574	0.920	-1.280	2.345
$\sigma_{\alpha_j\beta_j}$	-0.276	-0.253	0.221	-0.764	0.071
α	0.757	0.755	0.809	-0.830	2.333
β	1.118	1.117	0.044	1.034	1.206

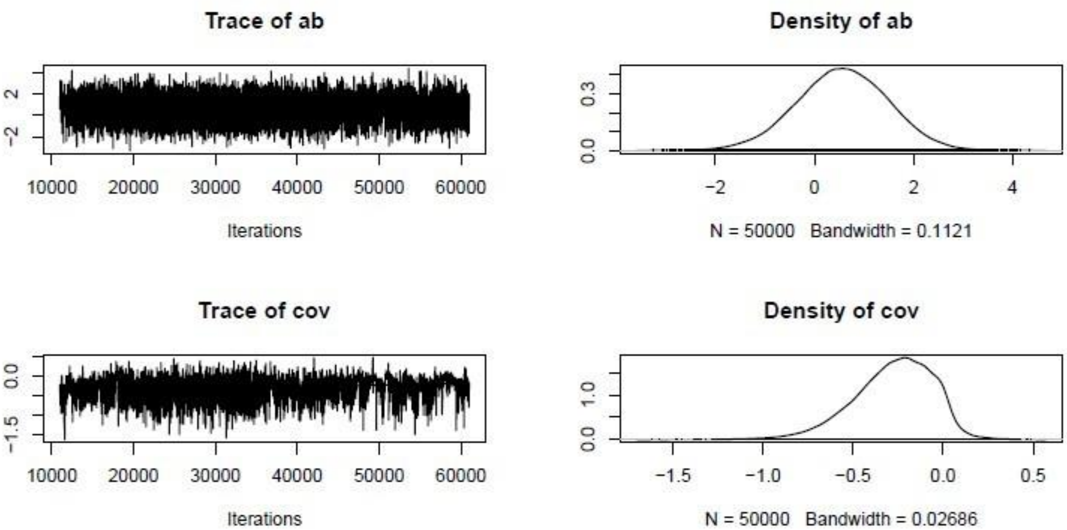

Figure 6.1 Trace plot and density plot for the posterior samples of the indirect effect (ab) and the covariance ($\sigma_{\alpha_j\beta_j}$) in the Bayesian partially clustered mediation model.

MODEL PERFORMANCE IN SIMULATION STUDY

SIMULATION DESCRIPTION

A simulation study was conducted to assess Bayesian estimation of the key parameter of interest in the multilevel mediation model for partially clustered design. The true values of the parameters were based on values used in three previous, related articles, including: Bauer, Preacher, and Gil (2006), Yuan and MacKinnon (2009), and Baldwin and Fellingham (2012) and were as follows:

$$\beta_1 \sim N\ (0, 60), \tag{7-1}$$

$$\beta_2 \sim N\ (0, 40), \tag{7-2}$$

$$\beta_{3j} \sim N\ (2, 4), \tag{7-3}$$

$$\sigma^2_{\alpha_j} = \sigma^2_{\beta_j} = 16. \tag{7-4}$$

Given the characteristics of the partially clustered design, heterogeneity across study conditions (treatment versus control arms) was assumed. The generating values used for level-1 residuals were as follows:

$$e_{1cluster_{ij}} \sim N\ (0, 65), \tag{7-5}$$

$$e_{1uncluster_{ij}} \sim N\ (0, 45), \tag{7-6}$$

$$e_{2cluster_{ij}} \sim N\ (0, 45), \tag{7-7}$$

$$e_{2uncluster_{ij}} \sim N\ (0, 26). \tag{7-8}$$

Note that the true values of the parameters proposed above were rescaled by multiplying the variance and covariance values by 100 and the original fixed parameter values by 10 (Baldwin & Fellingham, 2012; Yuan & MacKinnon, 2009). Based on previous work in this area, two design factors were manipulated to generate data fitting different scenarios. These design factors included the true values for α and β and the true covariance between α_j and β_j. The sample sizes at the first and second levels in this report were fixed at N_1=6 and N_2=100, where N1 and N2 denote the sample sizes at level-1 and level-2, respectively. Specifically, two sets of effect sizes (α=β=3 or α=β=6) and three values of covariance (−11.3, 0, 11.3) were studied. Overall, the two design factors above together yielded a total of 6 scenarios. One thousand data sets were generated for each scenario and the model estimated using Bayesian methods. The partially clustered multilevel mediation model discussed above was applied to each simulated dataset. We assessed estimation of the indirect effect and of the covariance between α_j and β_j. Parameter recovery was evaluated using the relative parameter bias as well as coverage rate for the 95% credible interval.

SIMULATION RESULTS

Again, the same noninformative priors were used as in the real data analysis, including a scaled inverse Wishart prior for the covariance matrix in the bivariate normal distribution of α_j and β_j was used (Gelman & Hill, 2007). Specifically, vague normal priors were assigned to the regression parameters:

$$\mathrm{p}(\beta_1) \sim N\,(0, 10^6), \tag{7-9}$$

$$\mathrm{p}(\beta_2) \sim N\,(0, 10^6), \tag{7-10}$$

$$p(\gamma_{30}) \sim N\,(0, 10^6), \tag{7-11}$$

$$p(\alpha_j, \beta_j) \sim Bivariate\ Normal\ (\binom{\alpha}{\beta}, \Sigma), \tag{7-12}$$

where the scaled inverse Wishart priors were specified with a degrees of freedom of 2 and a scale matrix of $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$ applied to the covariance matrix Σ as discussed in the real-data study, and

$$p(\alpha) \sim N\,(0, 10^6), \tag{7-13}$$

$$p(\beta) \sim N\,(0, 10^6). \tag{7-14}$$

Vague gamma priors were assigned for the level-1 variance parameters:

$$p(\sigma_1^{-2}) \sim Gamma\,(0.001, 0.001), \tag{7-15}$$

$$p(\sigma_2^{-2}) \sim Gamma\,(0.001, 0.001). \tag{7-16}$$

For the variance of β_{3j}, a vague uniform prior was used:

$$p(\sigma_{\beta 3j}) \sim Unif\,(0,100), \tag{7-17}$$

where $\sigma_{\beta 3j}$ represents the standard deviation of β_{3j}. 60,000 posterior draws are used to make Bayesian inferences with the first 20,000 as burn-in iterations. Results focus on inferences of average indirect effect and the covariance between α_j and β_j and are shown in Table 7.1.

When the sample size of the first-level units and of the second-level units were fixed at 6 and 100, respectively, the Bayesian point estimates of the indirect effects as well as the covariance between α_j and β_j showed negligible bias except in the scenario with a small effect size and a negative covariance between α_j and β_j. The coverage rates of the 95% credible interval were also shown to be quite close to the nominal level.

Overall, no matter whether the scale of the mediation effect was small or moderate and whether the covariance between α_j and β_j was positive, negative or zero, when sample sizes were moderately large, the results indicated good parameter recovery for the proposed partially clustered multilevel mediation model.

Table 7.1 Relative Parameter Bias and Coverage Rates (%) for the 95% Credible Intervals for the Indirect Effects and Covariance Estimates by Condition

	α=β=3			**α=β=6**		
	$\boldsymbol{\sigma_{\alpha_j\beta_j} = -11.3}$	$\boldsymbol{\sigma_{\alpha_j\beta_j} = 0}$	$\boldsymbol{\sigma_{\alpha_j\beta_j} = 11.3}$	$\boldsymbol{\sigma_{\alpha_j\beta_j} = -11.3}$	$\boldsymbol{\sigma_{\alpha_j\beta_j} = 0}$	$\boldsymbol{\sigma_{\alpha_j\beta_j} = 11.3}$
Bias (*ab*)	0.081	-0.005	0.001	-0.012	-0.011	-0.014
Coverage Rate (*ab)*	94.9%	94.4%	94.0%	95.4%	94.6%	92.1%
Bias (*cov*)	-0.0003	-0.055*	-0.015	-0.001	-0.037*	-0.021
Coverage Rate (*cov*)	93.7%	93.1%	93.7%	93.8%	93.8%	93.3%

* When the true value of the covariance is set at zero, parameter bias for the Bayesian estimates of covariance was reported instead.

DISCUSSION

The partially clustered design with a nonparallel structure is increasingly common in applied research. An appropriate statistical model that is designed to evaluate intervention effects in partially clustered data has been recently suggested (Baldwin, Bauer, Stice & Rohde, 2011) and extended in the current study for handling tests of mediation. Likelihood-based estimation procedures uses an adjustment to the standard errors of fixed effects to address the uncertainties associated with estimating the variance and covariance components through inflation of the standard errors. More specifically, the sampling distribution for the fixed effects are approximated by the *t*-distribution, with degrees of freedom approximated by the Satterthwaite adjustment. However, these adjustments turn out not to work well when sample sizes are small, resulting in variance component estimates at the boundary (Kenward & Roger, 1997; Baldwin & Fellingham, 2012). Instead, use of a Bayesian approach could provide a useful alternative that should better handle some of these complications.

In this paper, a Bayesian multilevel mediation model in the context of partially clustered design data is proposed to examine a mediation effect. Compared with the conventional likelihood-based approach, the Bayesian approach has several attractive features. First, use of prior distributions allows historical information from literature or pilot studies to be incorporated into the current data analysis. Second, Bayesian inference does not rely on large sample approximations (Robert, 2007). The posterior distribution that combines information from both the data and the prior distribution truly represents how the parameters might be distributed. Third, through the use of prior information, Bayesian methods automatically allows for uncertainties in the parameters without requiring any additional adjustments, and also avoids the boundary problems that are met in maximum likelihood estimation (Baldwin & Fellingham, 2012).

Our results from the simulation study suggest that the proposed Bayesian partially clustered mediation model works well regardless of the true values of the mediation effects as well as of the covariance between α_j and β_j, in larger sample size (N1=6, N2=100) conditions . However, only a limited set of conditions were examined here. The influence of sample size per se was not examined explicitly in this report. In Yuan and MacKinnon (2009), the authors found that the coverage rates for the 95% credible interval of the Bayesian estimates of the indirect effects and covariance between α_j and β_j were sensitive to sample size. When sample sizes were small, coverage rates tended to be less than nominal rates; and as sample size increased, coverage rates started better approximating nominal values.

The current study also did not examine a related question: is model performance sensitive to the choices of prior distribution or not? As we discussed in previous chapters, the choices of appropriate priors especially in small-sample situations is important as the prior plays a larger role in derivation of the posterior distributions. The use of inappropriate priors can result in biased estimates when using a Bayesian model. When sample sizes are small, non-informative diffuse priors might be unnecessarily inefficient, as their use implies that extreme values are as likely as non-extreme values (Baldwin & Fellingham, 2012). Therefore, it is necessary to elicit appropriate (informative) priors as summaries of the prior information, and possibly compare model fit under different prior distributions' specifications. Sensitivity analyses focused on the effect of different prior distributions' specifications could be explored. Last, only a subset of possible conditions was explored in the current study. Future research should extend these conditions further to better validate Bayesian estimation of the proposed multilevel mediation model for partially clustered data.

REFERENCE

Baldwin, S. A., Bauer, D. J., Stice, E., & Rohde, P. (2011). Evaluating Models for Partially Clustered Designs. *Psychological Methods, 16*(2), 149-165. doi: 10.1037/a0023464

Baldwin, S. A., & Fellingham ,G.W.(2012). Bayesian Methods for the Analysis of Small Sample Multilevel Data With a Complex Variance Structure.Psychological Methods, No Pagination Specified. doi:10.1037/a0030642

Bauer, D. J., Preacher, K. J., & Gil, K. M. (2006). Conceptualizing and testing random indirect effects and moderated mediation in multilevel models: New procedures and recommendations. *Psychological Methods*, *11*(2), 142-163. doi: 10.1037/1082-989x.11.2.142

Bauer, D. J., Sterba, S. K., & Hallfors, D. D. (2008). Evaluating group-based interventions when control participants are ungrouped. *Multivariate Behavioral Research, 43*(2), 210-236. doi: 10.1080/00273170802034810

Bollen, K. A., & Stine, R. (1990). Direct and indirect effects: Classical and bootstrap estimates of variability. *Sociological Methodology, 20,* 115–140.

Daniels, M. J., Roy, J. A., Kim, C., Hogan, J. W., & Perri, M. G. (2012). Bayesian Inference for the Causal Effect of Mediation. *Biometrics, 68*(4), 1028-1036. doi: 10.1111/j.1541-0420.2012.01781.x

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (2003). *Bayesian data analysis* (2nd ed.). New York, NY: Chapman & Hall/CRC.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York, NY: Cambridge University Press.

Geweke, J. (1992), Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments. Bayesian Statistics 4, Oxford: Oxford University Press, 169-193.

Hoover, D. R. (2002). Clinical trials of behavioural interventions with heterogeneous teaching subgroup effects. *Statistics in Medicine, 21*(10), 1351-1364. doi: 10.1002/sim.1139

Kenward, M. G., & Roger, J. H. (1997). Small sample inference for fixed effects from restricted maximum likelihood. *Biometrics, 53*(3), 983-997. doi: 10.2307/2533558

Kenward, M. G., & Roger, J. H. (2009). An improved approximation to the precision of fixed effects from restricted maximum likelihood. *Computational Statistics & Data Analysis, 53*(7), 2583-2595. doi: 10.1016/j.csda.2008.12.013

Krull, J. L., & MacKinnon, D. P. (2001). Multilevel modeling of individual and group level mediated effects. *Multivariate Behavioral Research, 36*(2), 249-277. doi: 10.1207/s15327906mbr3602_06

Lee, K. J., & Thompson, S. G. (2005). The use of random effects models to allow for clustering in individually randomized trials. *Clinical Trials, 2*(2), 163-173. doi: 10.1191/1740774505cn082oa

Mackinnon, D. P., & Dwyer, J. H. (1993). ESTIMATING MEDIATED EFFECTS IN PREVENTION STUDIES. *Evaluation Review, 17*(2), 144-158. doi: 10.1177/0193841x9301700202

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58*, 593-614. doi: 10.1146/annurev.psych.58.110405.085542

MacKinnon, D. P., Lockwood, C. M., Hoffman, J. M., West, S. G., & Sheets, V. (2002). A comparison of methods to test mediation and other intervening variable effects. *Psychological Methods, 7*(1), 83-104. doi: 10.1037//1082-989x.7.1.83

MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research, 39*(1), 99-128. doi: 10.1207/s15327906mbr3901_4

Myers, J. L., Dicecco, J. V., & Lorch, R. F. (1981). GROUP-DYNAMICS AND INDIVIDUAL-PERFORMANCES - PSEUDOGROUP AND QUASI-F ANALYSES. *Journal of Personality and Social Psychology, 40*(1), 86-98. doi: 10.1037//0022-3514.40.1.86

Pituch, K. A., & Stapleton, L. M. (2008). The performance of methods to test upper-level mediation in the presence of nonnormal data. *Multivariate Behavioral Research, 43*(2), 237-267. doi: 10.1080/00273170802034844

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods Instruments & Computers, 36*(4), 717-731. doi: 10.3758/bf03206553

Roberts, C., & Roberts, S. A. (2005). Design and analysis of clinical trials with clustering effects due to treatment. *Clinical Trials, 2*(2), 152-162. doi:

10.1191/1740774505cn076oa

Robert, C.P. (2007). *The Bayesian choice: From decision-theoretic foundations to computational implementation* (2nd ed.). New York, NY: Springer Verlag.

Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods, 7*(4), 422-445. doi: 10.1037//1082-989x.7.4.422

Yuan, Y., & MacKinnon, D. P. (2009). Bayesian Mediation Analysis. *Psychological Methods, 14*(4), 301-322. doi: 10.1037/a0016972